Here I define the Bayesian model I want to use to test the performance of the different priors. The features I want to include are the folowing: continuous outcome, any number/measurement scale.

Let's start with the **model**

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}^T\boldsymbol{\theta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \epsilon_{ij} \\ \boldsymbol{b}_i &\sim N(\boldsymbol{0}, \boldsymbol{\Psi}) \\ \epsilon_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{1}$$

(Vectors are in bold, matrix are capital Greek letters).
I'm going to define the following **priors**:

$$p(\boldsymbol{\theta}) \propto 1 \tag{2}$$

$$p(\sigma^2) \propto \sigma^{-2} \tag{3}$$

For what concerns the random effects variance covariance matrix, different priors are tested. In particular we used:

- inverse-Wishart

$$p(\boldsymbol{\Psi}) \propto IW(\nu, S_0) \tag{4}$$

where we choose $\nu = k - 1 + e$, and $S_0 = diag(k - 1 + e)$, following indications by Gelman *et al* (2014). Thanks to the partitioning property, we know that the diagonal components of Ψ have an inverse-Wisharet distribution themselves. Furthermore, for a univariate case, we know that an inverse Wishart distribution simplifies to an inverse Gamma with parameters $k = 1, \alpha = \frac{\nu}{2}, \beta = \frac{S_{0kk}}{2}$. The inverse-Wishart priors we are trying are:

Prior Description	ν	S_0
1. IW uninformative	$k - 1 + e$	diag(k-1+e)
2. IW educated	$k - 1 + e$	educated guess

- inverse-Wishart *a là* Huang and Wand

$$\begin{aligned} p(\boldsymbol{\Psi}|a_1, a_2) &\propto IW(\nu + k - 1, 2\nu \times diag(1/a_1, 1/a_2)), \\ a_k &\propto IG(1/2, 1/A_k^2), \end{aligned} \tag{5}$$

with $\nu = 2$ and $\boldsymbol{A} = [1000, 1000]$. The marginal distribution of any standard deviation term in $\boldsymbol{\Psi}$ is Half-$t(\nu, A_k)$ and, when choosing $\nu = 2$, the marginal disitbrution on the correlation term is uniform on (-1, 1), see property 2 to 4 in Huang and Wand (2013, p. 442). Furthermore, according to Huang and Wand (2013, p. 441) arbitrarily large values for a_k lead to arbitrarily weak priors on the standard deviation term. Hence, our choices for the parameter of this prior are:

Prior Description	ν	$\boldsymbol{A}$
3. IW a là HW	2	$[1000, 1000]$

- Matrix-F variate

$$\begin{aligned} p(\boldsymbol{\Psi}) &\propto F(\boldsymbol{\Psi}; \nu, \delta, \boldsymbol{B}) \\ &\propto \int IW(\boldsymbol{\Psi}; \delta + k - 1, \Sigma) \times W(\boldsymbol{\Sigma}; \nu, \boldsymbol{B}) d\boldsymbol{\Sigma} \end{aligned} \tag{6}$$

with degrees of freedom $\nu > k-1$, $\delta > 0$, and $\boldsymbol{B}$ a positive definite scale matrix that functions as prior guess. Three different choices where made for $\boldsymbol{B}$ in this paper: diag(10^3), proper neighbor of $(\sigma^2)^{-\frac{1}{2}}$; $\boldsymbol{B}_{ed}$, an educated guess based on data exploration, $\boldsymbol{R^*}$ and an empirical bayes choice following Kass and Natarajan (2006).
Considering a 2×2 random effects variance covariance matrix (random intercepts, and random slopes) that is matrix-F distirbuted, $F(\nu,\delta,\boldsymbol{B})$, the marginal distribution on the standard deviations of the random effects are univariate $F(\nu,\delta,b_{11})$ and $F(\nu,\delta,b_{22})$, with $\nu > 1, \delta > 0, b_{jj} > 0$. There we chose the first integer number we could for the parameters ν, and δ. When I chose $\delta = 2$, I was thinking of staying close to what has been done in Mulder Pericchi. When I chose $\delta = e$ I wanted to try with some uninformative prior that got as close as possible to non-informative.

Prior Description	ν	δ	S_0
4. mat-F proper neighbor	2	2	$10^3 \times \boldsymbol{I}_2$
5. mat-F uninformative	$k-1+e$	e	educated guess
6. mat-F educated guess	2	2	educated guess
7. mat-F empirical Bayes	2	2	$\boldsymbol{R}^*$

The derivation of the conditional posterior follows.

Full conditional for $\boldsymbol{\theta}$ (fixed effects)

Let's start with

$$p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{b}_i,\boldsymbol{\Psi},\sigma^2) = p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{b}_i,\boldsymbol{\Psi},\sigma^2)p(\boldsymbol{\theta})$$

where

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Psi},\sigma^2) &= \prod_{i=1}^{n}\prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\theta}^T\boldsymbol{x}_{ij} + \boldsymbol{b}_i^T\boldsymbol{z}_{ij},\boldsymbol{\Psi},\sigma^2) \\ &\propto exp(-\frac{1}{2\sigma^2}SSR) \end{aligned}$$

and

$$SSR = \sum_{i=1}^{n}\sum_{j=1}^{J}(y_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij} - \boldsymbol{b}_i^T\boldsymbol{z}_{ij})^2$$

where can rewrite y_{ij} as $\tilde{y}_{ij}$, with $\tilde{y}_{ij} = y_{ij} - \boldsymbol{b}_i^T\boldsymbol{z}_{ij}$ which makes SSR:

$$\begin{aligned} SSR &= \sum_{i=1}^{n}\sum_{j=1}^{J}(\tilde{y}_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij})^2 \\ &= (\tilde{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\theta})^T(\tilde{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\theta}) \\ &= \tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}} - 2\boldsymbol{\theta}^T\boldsymbol{X}\tilde{\boldsymbol{y}} + \boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} \end{aligned}$$

Hence,

$$p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Psi},\sigma^2) \propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{\theta}^T\boldsymbol{X}\tilde{\boldsymbol{y}} + \boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta}])$$

Combining this with the prior we obtain:

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\Psi},\sigma^2) &\propto exp(-\frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\theta}^T\boldsymbol{X}\tilde{\boldsymbol{y}}) \\ \boldsymbol{\theta}|. &\sim \boldsymbol{N}\left(\frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}\tilde{\boldsymbol{y}}}{\sigma^2}, \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}}{\sigma^2}\right) \end{aligned} \tag{7}$$

Full conditional for $\boldsymbol{b}_i$ (random effects)

To derive this one we can start from:

$$p(\boldsymbol{b}_i|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\Psi}, \sigma^2) = p(\boldsymbol{y_i}|\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2)p(\boldsymbol{b}_i)$$

We know that

$$p(\boldsymbol{y}_i|.) = \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\theta}^T\boldsymbol{x}_{ij} + \boldsymbol{b}_i^T\boldsymbol{z}_{ij}, \sigma^2) \propto exp(-\frac{1}{2\sigma^2}SSR_i)$$

with

$$SSR = \sum_{j=1}^{J}(y_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij} - \boldsymbol{b}_i^T\boldsymbol{z}_{ij})^2$$

and we can rewrite y_{ij} as $\tilde{y}_{ij} = y_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij}$, which would make SSR be

$$\begin{aligned}
SSR &= \sum_{j=1}^{J}(\tilde{y}_j - \boldsymbol{\theta}^T\boldsymbol{x}_j)^2 \\
&= (\tilde{\boldsymbol{y}} - \boldsymbol{b}_i^T\boldsymbol{Z}_i)^T(\tilde{\boldsymbol{y}} - \boldsymbol{b}_i^T\boldsymbol{Z}_i) \\
&= \tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}} - 2\boldsymbol{b}_i^T\boldsymbol{Z}_j\tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T\boldsymbol{Z}_i^T\boldsymbol{Z}_j\boldsymbol{b}_i
\end{aligned}$$

Hence,

$$p(\boldsymbol{y}_i|.) \propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T\boldsymbol{Z}_j\tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T\boldsymbol{Z}_i^T\boldsymbol{Z}_j\boldsymbol{b}_i])$$

We also know that in this case, the "prior" is

$$p(\boldsymbol{b}_i) \propto N(\boldsymbol{0}, \boldsymbol{\Psi}) \propto exp(-\frac{1}{2}[-2\boldsymbol{b}_i^T\boldsymbol{\Psi}^{-1}\boldsymbol{0} + \boldsymbol{b}_i^T\boldsymbol{\Psi}^{-1}\boldsymbol{b}_i])$$

In conclusion, combining the sampling model and the prior, we get:

$$\begin{aligned}
p(\boldsymbol{b}_i|.) &\propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T\boldsymbol{Z}_j\tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T\boldsymbol{Z}_i^T\boldsymbol{Z}_j\boldsymbol{b}_i] - \frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T\boldsymbol{Z}_j\tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T\boldsymbol{Z}_i^T\boldsymbol{Z}_j\boldsymbol{b}_i]) \\
\boldsymbol{b}_i|. &\propto \boldsymbol{N}\left(\left(\Psi^{-1} + \frac{\boldsymbol{Z}_i^T\boldsymbol{Z}_i}{\sigma^2}\right)^{-1}\left(\boldsymbol{\Psi}^{-1}\boldsymbol{0} + \frac{\boldsymbol{Z}_i^T\tilde{y}_i}{\sigma^2}\right), \left(\Psi^{-1} + \frac{\boldsymbol{Z}_i^T\boldsymbol{Z}_i}{\sigma^2}\right)^{-1}\right)
\end{aligned} \tag{8}$$

Full conditional for σ^2 (error variance)

The full conditional posterior can be expressed as:

$$p(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{\Psi}) = p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2)p(\sigma^2)$$

The sampling model is the same we saw for the full conditional distribution of $\boldsymbol{\theta}$:

$$\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) &= \prod_{i=1}^{n}\prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\theta}^T\boldsymbol{x}_{ij} + \boldsymbol{b}_i^T\boldsymbol{z}_{ij}, \boldsymbol{\Psi}, \sigma^2) \\
&= \prod_{i=1}^{n}\prod_{j=1}^{J}(2\pi\sigma^{-2})^{-\frac{1}{2}}exp(-\frac{(y_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij} - \boldsymbol{b}_i^T\boldsymbol{z}_{ij})^2}{2\sigma^2})
\end{aligned}$$

However, we are now interested in σ^2, hence

$$\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) &\propto (\sigma^2)^{-\frac{N}{2}}exp(-\frac{\sum_{i=1}^{n}\sum_{j=1}^{J}(y_{ij} - \boldsymbol{\theta}^T\boldsymbol{x}_{ij} - \boldsymbol{b}_i^T\boldsymbol{z}_{ij})^2}{2\sigma^2}) \\
&\propto (\sigma^2)^{-\frac{N}{2}}exp(-\frac{1}{2\sigma^2}SSR)
\end{aligned}$$

where $N = \sum_i^n nj_i$ is the entire sample size (all observations within all clusters). The prior for σ is given above, and therefore we can write the full conditional posterior as:

$$\begin{aligned} p(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{\Psi}) &\propto (\sigma^2)^{-\frac{N}{2}-1} exp(-\frac{1}{2\sigma^2} SSR) \\ \sigma^2|. &\sim IG(\frac{N}{2}, \frac{SSR}{2}) \end{aligned} \tag{9}$$

Full conditional for $\boldsymbol{\Psi}$ (random effects variance covariance matrix)

Here, we need to write down the posteriors for the different priors we specified. First, let us define the sampling model for the random effects.

$$\begin{aligned} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} &= \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Psi}) \\ p(\boldsymbol{b}_1, \boldsymbol{b}_2|\boldsymbol{\Psi}) &\propto |\boldsymbol{\Psi}|^{-\frac{n}{2}} exp\left(-\frac{1}{2} tr(\boldsymbol{S}_b \boldsymbol{\Psi}^{-1})\right) \end{aligned} \tag{10}$$

where $\boldsymbol{S}_b$ is $\Sigma_i \boldsymbol{b}_i \boldsymbol{b}_i^T$

- given the inverse-Wishart prior

$$\begin{aligned} p(\boldsymbol{\Psi}) &\propto IW(\nu, \boldsymbol{S}_0) \\ &\propto |\boldsymbol{\Psi}|^{-\frac{(\nu+k+1)}{2}} exp\left(-\frac{1}{2} tr(\boldsymbol{S}_0 \boldsymbol{\Psi}^{-1})\right) \end{aligned}$$

  the full conditional posterior of $\boldsymbol{\Psi}$ is

$$\begin{aligned} p(\boldsymbol{\Psi}|.) &\propto |\boldsymbol{\Psi}|^{-\frac{(\nu+n+k+1)}{2}} exp\left(-\frac{1}{2} tr([\boldsymbol{S}_0 + \boldsymbol{S}_b] \boldsymbol{\Psi}^{-1})\right) \\ &\propto IW(\nu + n, \boldsymbol{S}_0 + \boldsymbol{S}_b) \end{aligned} \tag{11}$$

  where $\nu = 2$

- inverse-Wishart *a là* Huang and Wand

$$\begin{aligned} p(\boldsymbol{\Psi}|a_1, a_2) &\propto IW(\nu + k - 1, 2\nu diag(1/a_1, 1/a_2)), \\ a_k &\propto IG(\eta, 1/A_k^2) \\ p(\boldsymbol{\Psi}) &\propto |\boldsymbol{\Psi}|^{-\frac{(\nu+k-1+1)}{2}} exp\left(-\frac{1}{2} tr(2\nu diag(1/a_1, 1/a_2) \boldsymbol{\Psi}^{-1})\right) \\ &\times \left(\frac{1}{a_1}\right)^{\eta+1} exp\left(-\frac{1}{A_1^2 a_1}\right) \times \left(\frac{1}{a_2}\right)^{\eta+1} exp\left(-\frac{1}{A_2^2 a_2}\right) \end{aligned}$$

  the full conditional posterior of $\boldsymbol{\Psi}$ is

$$\begin{aligned} p(\boldsymbol{\Psi}|.) &\propto |\boldsymbol{\Psi}|^{-\frac{(\nu+k-1+n+1)}{2}} exp\left(-\frac{1}{2} tr([\boldsymbol{S}_b + 2\nu diag(1/a_1, 1/a_2)] \boldsymbol{\Psi}^{-1})\right) \\ &\propto IW(\nu + k - 1 + n, \boldsymbol{S}_b + 2\nu diag(1/a_1, 1/a_2)) \\ p(a_k|.) &\propto IG\left(\eta(\nu + k), \nu\left(\boldsymbol{\Psi}_{kk}^{-1} + \frac{1}{A_k^2}\right)\right) \end{aligned} \tag{12}$$

  where $\eta = \frac{1}{2}, \nu = 2, k = 2$, and n is the number of clusters (individuals). (For the conditional posterior of a_k refer to Huang and Wand (2013), section 4.2).

- Matrix-F variate

  Following section 2.3 in Mulder and Pericchi (2018), instead of working directly with the $\boldsymbol{\Psi} \sim F(\nu, \delta, \boldsymbol{B})$ we apply the parameter expansion defined above (see section on priors) and model it as $\boldsymbol{\Psi} \sim IW(\delta+k-1, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} \sim W(\nu, \boldsymbol{B})$. With this parameter expansion, the conditional priors are:

$$\boldsymbol{\Psi}|\boldsymbol{\Omega} \sim IW(\delta + k - 1, \boldsymbol{\Omega})$$
$$\boldsymbol{\Omega}|\boldsymbol{\Psi} \sim W(\nu + \delta + k - 1, (\boldsymbol{\Psi}^{-1} + \boldsymbol{B}^{-1})^{-1})$$

  which makes the full conditional posterior of:

$$\boldsymbol{\Psi}|\boldsymbol{\Omega}, . \sim IW(\delta + k - 1 + n, \boldsymbol{S}_b + \boldsymbol{\Omega})$$
$$\boldsymbol{\Omega}|\boldsymbol{\Psi}, . \sim W(\nu + \delta + k - 1, (\boldsymbol{\Psi}^{-1} + \boldsymbol{B}^{-1})^{-1})$$

  with parameters as defined above. Given these posteriors, the Gibbs sampler implementation is straightforward.

Notation Conventions

- n number of clusters; i specific cluster
- J number of observations within cluster; j specific observation
- N total number of observations